Supplementary Online Material for

A Cellular Solution to an Information Processing Problem

Garud Iyengar[1,*] and Madan Rao[2,3,*]

[1]Industrial Engineering and Operations Research, Columbia University, New York, USA

[2]Raman Research Institute, C.V. Raman Avenue, Bangalore 560080, India

[3]National Centre for Biological Sciences (TIFR), Bellary Road, Bangalore 560065, India

[*]To whom correspondence should be addressed; E-mail: garud@ieor.columbia.edu, madan@ncbs.res.in.

Contents

S1 Idealized model: analytical details

Here we attempt to understand the optimal sensor organization when transport is not limited by physical constraints. This idealized mathematical situation allows us to make accurate analytical statements which will be a useful guide to the more "realistic" stochastic model. The input signal f is taken to be piece-wise constant in time with $f(x,t) = f_k(x)$, for $t \in \big[k\tau, (k+1)\tau\big)$, where each function $f_k(x)$ is chosen from the following set of functions

$$\mathcal{F} = \left\{ f(x) : f(x) \in C^1, \left\| \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{bmatrix} \right\| \leq \frac{1}{\xi} \right\}, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm and ξ is the spatial correlation length of the signal[1].

We compare the performance of two different signal acquisition architectures - *stationary* and *mobile*. In the *perfect mobile* architecture all sensors move in a coordinated fashion with velocity v and in the *diffusion-advection mobile* architecture the sensor movement approximates diffusion-advection transport. We show that there is a phase transition from stationary to mobile architecture as a function of the sensor density ρ, sensor velocity v, sensor sampling time t_m, and the correlation length and time ξ and τ.

S1.1 Phase transition between stationary and perfect mobile architectures

Let $\mathcal{Z} = \{y : s(y) > 0\}$ denote the sensor locations, and suppose that the sensors sample the function every t_m seconds. Since the sensors take measurement at a frequency $\frac{1}{t_m}$, there are $\frac{\tau}{t_m}$ sampling instants in each signal period; consequently, the variance of the function estimate $\hat{f}(y)$ is given by

$$\mathbf{var}(\hat{f}(y) - f(\mathbf{y})) = \frac{\sigma_p^2}{(s(y))^{(1+\gamma)}} \cdot \frac{t_m}{\tau}.$$

We assume that the approximation $\hat{f}(x)$ at a location x with no sensors, i.e. $x \notin \mathcal{Z}$, function is reconstructed using a nearest neighbor mapping; i.e. the function estimate $\hat{f}(x)$ at a the location x is mapped to the location $y \in \mathcal{Z}$ that minimizes the total expected error

$$\epsilon(x|y) = \max_{f \in \mathcal{F}} \mathbb{E}(f(x) - \hat{f}(y))^2 = \max_{f \in \mathcal{F}} \mathbb{E}(f(x) - f(y))^2 + \frac{\sigma_p^2 t_m}{s(y)} = \frac{\|x-y\|^2}{\xi^2} + \frac{\sigma_p^2 t_m}{\big(s(y)\big)^{(1+\gamma)}},$$

where the last expression follows from the fact $f(x) - f(y) = \boldsymbol{\nabla} f(z)^\top (x - y)$ for some $z \in [x, y]$, and $\|\boldsymbol{\nabla} f(z)\| \leq \frac{1}{\xi}$ for all $f \in \mathcal{F}$. Therefore, it follows that the minimum mean-square error $\epsilon(x|s)$ at location x is given by[2]

$$\epsilon(x|s) = \min \left\{ \frac{\|x-y\|^2}{\xi^2} + \frac{\sigma_p^2 t_m}{\tau(s(y))^{(1+\gamma)}} : y \in \mathcal{Z} \right\}.$$

[1]The results do not change if the functions $f_k(x)$ are drawn from a Gaussian random field with correlations $\mathbf{cor}\big(f_k(x), f_l(y)\big) = \delta_{kl} K\left(\frac{\|x-y\|}{\xi}\right)$, where $K(\cdot)$ is a non-negative decreasing function of its argument.

[2]For the Gaussian random field f_k, $\epsilon(x|s) = \min \left\{ 2\sigma_f^2 \left(1 - K\left(\frac{\|x-y\|}{\xi}\right)\right) + \frac{\sigma_p^2 t_m}{\tau(s(y))^{(1+\gamma)}} : y \in \mathcal{Z} \right\}.$

The error $\epsilon(x|s)$ has two components. The first component is the spatial error associated with function de-correlation over the correlation length ξ, and the second component is the error associated with sensor noise. The worst-case overall estimation error $\epsilon(s) = \max_x\{\epsilon(x|s)\}$. We propose that the optimal sensor architecture minimizes the error $\epsilon(s)$. We use the maximum error criterion in order to avoid having to postulate a measure over the spatial and temporal dimension. The scaling results in this paper remain true if the maximum error is replaced by average error. Let the *Voronoi cell* $\mathcal{V}(y) = \{x : \mathrm{argmin}_{w\in\mathcal{Z}}\,\epsilon(x|w) = y\}$ denote the set of positions x that are mapped to a particular sensor location y. Then $\epsilon(s) = \max_{y\in Z}\max_{x\in\mathcal{V}(y)}\epsilon(x|y)$. Thus, the optimal sensor placement s is a solution to the optimization problem

$$\begin{array}{ll} \min & \max_{y\in Z}\max_{x\in\mathcal{V}(y)}\left\{\frac{\|x-y\|^2}{\xi^2} + \frac{\sigma_p^2 t_m}{\tau(s(y))^{(1+\gamma)}}\right\}, \\ \text{s.t.} & \sum_{y:y\in\mathcal{Z}} s(y) = \rho L^2, \quad s(y) \text{ non-negative integer.} \end{array} \tag{2}$$

The optimization problem (2) is a variant of vector quantization (*1*) where, in addition to the the sensor locations $\mathcal{Z}$, the number of sensors s at each location is also a decision variable. In the rest of the analysis, we will take the thermodynamic limit and relax the constraint that $s(y)$ is an integer; and replace it with the constraint, that either $s(y) = 0$, i.e. y is not a signal acquisition location, or $s(y) \geq 1$, i.e. there is at least one sensor at each signal acquisition location y.

In the limit of large L and t_m, an optimal solution is to tesselate space with identical Voronoi cells $\mathcal{V}$, and the set $\mathcal{Z}$ consists of the centroids of $\mathcal{V}$. The signal de-correlation error then depends on the distance $\|x-y\|$ from a measurement location y; therefore, it follows that one of the ideal Voronoi tesselation corresponds to the hexagonal packing with regular hexagons $\mathcal{H}(r)$ with a radius r that is function of the correlation length ξ, the density γ and γ.

Suppose $\mathcal{V} = \mathcal{H}(r)$. Since the area of the hexagon $\mathcal{H}(r) = \frac{3\sqrt{3}}{2}r^2$, it follows that $s = \frac{3\sqrt{3}}{2}(\rho r^2)$. Therefore, the maximum error $\epsilon_s(r)$ in the Voronoi cell $\mathcal{V}$ is given by

$$\epsilon_s(r) = \frac{r^2}{\xi^2} + \frac{2\sigma_p^2 t_m}{3\sqrt{3}\tau(\rho r^2)^{(1+\gamma)}}$$

It is convenient to use dimensionless quantities, $\ell = \frac{r}{\xi}$, $\delta = \frac{2\sigma_p^2}{3\sqrt{3}}\cdot\frac{t_m}{\tau}$, and $\lambda = \frac{1}{\sqrt{\rho\xi^2}}$. The error of stationary

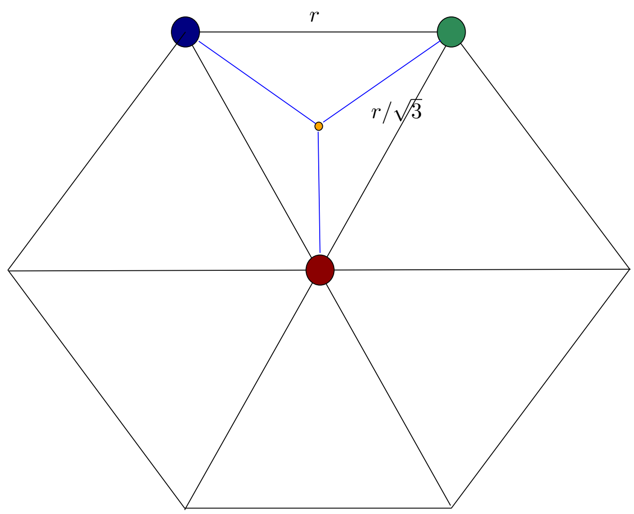


Figure S1: Perfect mobile architecture

architecture for a given radius ℓ is given by

$$\epsilon_s(\ell; \delta, \lambda) = \ell^2 + \frac{\delta \lambda^{2(1+\gamma)}}{\ell^{2(1+\gamma)}}. \tag{3}$$

The constraint $s(y) \geq 1$ translates to $\rho \xi^2 \ell^2 \geq 1$, or equivalently, $\ell \geq \lambda$. Thus, the optimal error of the stationary architecture is given by

$$\epsilon_s^*(\delta, \lambda) = \min_{\ell \geq \lambda} \{\epsilon_s(\ell; \delta, \lambda)\}.$$

Recall that the function estimation error is the sum of the spatial error and statistical error. The stationary mechanism is clearly optimal for minimizing the statistical error term since it maximizes the number of independent measurements at each sampling location y. It is possible that mobile sensors, that are able to sample the function at more locations, can significantly reduce the spatial error. However, mobility reduces the number of measurements at any given location increasing the statistical error.

Suppose in the hexagonal packing with radius r, instead of remaining stationary, the sensors alternate between the three different hexagonal packing configuration that all tesselate the plane [3]. Then the distance of any location x to a signal acquisition position y is at most $\frac{r}{\sqrt{3}}$ (see Figure S1), i.e. the spatial error term

[3]This analysis can be extended to a mobile architecture that alternates between an arbitrary number of configurations.

reduces to $(\frac{r}{\sqrt{3}\xi})^2$. However, the number of independent *temporal* samples at each sensor location is not $\frac{\tau}{t_m}$. Since the sensors move a distance r in every move, the re-organization of the sensors takes $\frac{r}{v}$ time units, the time available for sensing is $\frac{\tau}{3} - \frac{r}{v}$. Therefore, the sum of the spatial and the statistical for this architecture (in terms of the scaled distance ℓ) is given by

$$\epsilon_m(\ell; \delta, \lambda, \theta) = \frac{\ell^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{(1 - \frac{\ell}{\theta})\ell^{2(1+\gamma)}},$$

where $\theta = \frac{v\tau}{3\xi}$ is a dimension-less term that measures the trade-off between the correlation length ξ, velocity v and the signal duration (correlation time) τ.

The error $\epsilon_m(\ell; \delta, \lambda, \theta)$ is a decreasing function of θ; and, the minimum error $\epsilon_m^*(\delta, \lambda, \infty) = \min_{\ell \geq \lambda}\{\frac{\ell^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{\ell^{2(1+\gamma)}}\}$. Next, we compare $\epsilon_s^*(\delta, \lambda)$ and $\epsilon_m^*(\delta, \lambda, \infty)$ for different values of γ.

(a) $\gamma = 0$, i.e. independent sensors. We have to consider three cases:

(i) $\delta^{\frac{1}{4}}\lambda^{\frac{1}{2}} > \lambda$, i.e. $\lambda < \sqrt{\delta}$, or equivalently, $\rho\xi^2 \geq \frac{1}{\delta}$: In this case, $\ell_s^* = \delta^{\frac{1}{4}}\lambda^{\frac{1}{2}}$ and $\ell_m^* = \big(9\delta\big)^{\frac{1}{4}}\lambda^{\frac{1}{2}}$, and the optimal error of both architectures is

$$\epsilon_s^*(\delta, \lambda) = \epsilon_m^*(\delta, \lambda, \infty) = 2\lambda\sqrt{\delta},$$

i.e. the mobile and stationary architecture have the same optimal error.

(ii) $\sqrt{3}\delta^{\frac{1}{4}}\lambda^{\frac{1}{2}} \geq \lambda > \delta^{\frac{1}{4}}\lambda^{\frac{1}{2}}$, or equivalently, $\sqrt{\delta} < \lambda < 3\sqrt{\delta}$: In this case, In this case, $\ell_s^* = \lambda$ and $\ell_m^* = \big(9\delta\big)^{\frac{1}{4}}\lambda^{\frac{1}{2}}$, and the optimal error

$$\epsilon_s^*(\delta, \lambda) = \lambda^2 + \delta > \epsilon_m^*(\delta, \lambda, \infty) = 2\lambda\sqrt{\delta},$$

i.e. the mobile architecture has strictly superior performance.

(iii) $\sqrt{3}\delta^{\frac{1}{4}}\lambda^{\frac{1}{2}} < \lambda$, or equivalently, $\lambda > 3\sqrt{\delta}$: In this case, In this case, $\ell_s^* = \ell_m^* = \lambda$, and the optimal error

$$\epsilon_s^*(\delta, \lambda) - \epsilon_m^*(\delta, \lambda, \infty) = (\lambda^2 + \delta) - \left(\frac{\lambda^2}{3} + 3\delta\right) = \frac{2}{3}(\lambda^2 - 3\delta) > \frac{2}{3}(9\delta - 3\delta) = 4\delta > 0,$$

i.e. the mobile architecture has strictly superior performance.

(b) $\gamma > 0$, i.e. sensors display allosteric interaction. We have to consider three cases:

(i) $\lambda < \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, i.e. $\lambda < \sqrt{\delta(1+\gamma)}$: In this case, $\ell_s^* = \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$ and $\ell_m^* = \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, and the optimal error

$$\epsilon_m^*(\delta, \lambda, \infty) = 3^{-\frac{\gamma}{\gamma+2}} \epsilon_s^*(\delta, \lambda) = 3^{-\frac{\gamma}{\gamma+2}} \delta^{\frac{1}{\gamma+2}} \lambda^{\frac{2(\gamma+1)}{\gamma+2}} \left(\frac{\gamma+2}{(1+\gamma)^{\frac{2}{\gamma+2}}} \right),$$

i.e. mobile architecture has strictly superior performance.

(ii) $\big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}} < \lambda \le \big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, or equivalently, $\sqrt{(1+\gamma)\delta} < \lambda \le 3\sqrt{(1+\gamma)\delta}$: In this case, $\ell_s^* = \lambda$ and $\ell_m^* = \big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, and the optimal error

$$\epsilon_s^*(\delta, \lambda) = \lambda^2 + \delta > 3^{\frac{\gamma}{\gamma+2}} \epsilon_m^*(\delta, \lambda, \infty),$$

i.e. mobile architecture has strictly superior performance.

(iii) $\big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}} < \lambda$, or equivalently, $\lambda > 3\sqrt{(1+\gamma)\delta}$: In this case, In this case, $\ell_s^* = \ell_m^* = \lambda$, and the optimal error

$$\epsilon_s^*(\delta, \lambda) - \epsilon_m^*(\delta, \lambda, \infty) = (\lambda^2 + \delta) - \left(\frac{\lambda^2}{3} + 3\delta \right) = \frac{2}{3}(\lambda^2 - 3\delta) > \frac{2}{3}(9(1+\gamma)\delta - 3\delta) > 0,$$

i.e. mobile architecture has strictly superior performance.

(c) $-1 < \gamma < 0$, i.e. sensors display negative interaction. We have to consider three cases:

(i) $\lambda < \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, i.e. $\lambda < \sqrt{\delta(1+\gamma)}$: In this case, $\ell_s^* = \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$ and $\ell_m^* = \big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, and the optimal error

$$\epsilon_m^*(\delta, \lambda, \infty) = 3^{-\frac{\gamma}{\gamma+2}} \epsilon_s^*(\delta, \lambda) = 3^{-\frac{\gamma}{\gamma+2}} \delta^{\frac{1}{\gamma+2}} \lambda^{\frac{2(\gamma+1)}{\gamma+2}} \left(\frac{\gamma+2}{(1+\gamma)^{\frac{2}{\gamma+2}}} \right),$$

i.e. the stationary architecture has strictly superior performance.

(ii) $\big(\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}} < \lambda \le \big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$, or equivalently, $\sqrt{(1+\gamma)\delta} < \lambda \le 3\sqrt{(1+\gamma)\delta}$: In this case, $\ell_s^* = \lambda$ and $\ell_m^* = \big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}} \lambda^{\frac{\gamma+1}{\gamma+2}}$. In this parameter regime, the error of the stationary architecture is smaller (resp. larger) than the error of the mobile architecture, when $\lambda^2 + \delta$ is less than (resp. larger than) $3^{-\frac{\gamma}{\gamma+2}} \delta^{\frac{1}{\gamma+2}} \lambda^{\frac{2(\gamma+1)}{\gamma+2}} \left(\frac{\gamma+2}{(1+\gamma)^{\frac{2}{\gamma+2}}} \right)$.

(iii) $\big(9\delta(1+\gamma)\big)^{\frac{1}{2(\gamma+2)}}\lambda^{\frac{\gamma+1}{\gamma+2}} < \lambda$, or equivalently, $\lambda > 3\sqrt{(1+\gamma)\delta}$: In this case, In this case, $\ell_s^* = \ell_m^* = \lambda$, and the optimal error

$$\epsilon_s^*(\delta,\lambda) - \epsilon_m^*(\delta,\lambda,\infty) = (\lambda^2+\delta) - \left(\frac{\lambda^2}{3} + 3\delta\right) = \frac{2}{3}(\lambda^2 - 3\delta) > 0,$$

provided $\lambda^2 > 3\delta$, i.e. the mobile architecture has a strictly superior performance for large enough λ.

The results of this analysis are summarized in Table 1. Next, we analyze the performance for finite θ.

$$\epsilon_m^*(\eta,\theta) = \min_{\ell\geq\lambda}\left\{\frac{\ell^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{(1-\ell/\theta)\ell^{2(1+\gamma)}}\right\} = \theta^2\cdot\min_{x\geq\frac{\lambda}{\theta}}\left\{\frac{x^2}{3} + \frac{3\big(\eta(\frac{\lambda}{\theta})^{2(\gamma+2)}\big)}{(1-x)x^{2(1+\gamma)}}\right\}.$$

Let $f(y,z) = \min_{x\geq z}\left\{\frac{x^2}{3} + \frac{3y}{x^{2(1+\gamma)}}\right\}$. Then the phase transition between mobile and stationary architecture is given by

$$\theta^2 f\left(\frac{\delta\lambda^{2(1+\gamma)}}{\theta^{2(\gamma+2)}}, \frac{\lambda}{\theta}\right) \leq \epsilon_s^*(\delta,\lambda).$$

The function $f(y,z)$ is hard to compute, so we will attempt to approximate the phase boundary. Suppose $\theta \geq \alpha\ell_m^*(\delta,\lambda,\infty)$. Then

$$\begin{aligned}
\epsilon_m^*(\delta,\lambda,\theta) &= \min_{\ell\geq\lambda}\left\{\frac{\ell^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{(1-\ell/\theta)\ell^{2(1+\gamma)}}\right\}, \\
&\leq \frac{\big(\ell_m^*(\delta,\lambda,\infty)\big)^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{\big(1-\ell_m^*(\eta,\infty)/\theta\big)\big(\ell_m^*(\delta,\lambda,\infty)\big)^{2(1+\gamma)}} \\
&\leq \frac{\ell_m^*(\delta,\lambda,\infty)^2}{3} + \left(\frac{\alpha}{\alpha-1}\right)\frac{3\delta\lambda^{2(1+\gamma)}}{\big(\ell_m^*(\delta,\lambda,\infty)\big)^{2(1+\gamma)}}, \\
&\leq \left(\frac{\alpha}{\alpha-1}\right)\epsilon_m^*(\delta,\lambda,\infty).
\end{aligned}$$

Therefore, it follows that the error $\epsilon_m^*(\delta,\lambda,\theta) \leq \epsilon_s(\lambda,\theta)$ whenever $\theta \geq \alpha\ell_m^*(\delta,\lambda,\infty)$, and $\epsilon_m^*(\delta,\lambda,\infty) \leq \left(\frac{\alpha-1}{\alpha}\right)\epsilon_s^*(\delta,\lambda)$. Consider, for example, the parameter regime $\gamma \geq 0$, and $\lambda > 3\sqrt{(1+\gamma)\delta}$. In this regime $\ell_m^*(\delta,\lambda,\infty) = \lambda$, and $\frac{\epsilon_m^*(\delta,\lambda,\infty)}{\epsilon_s^*(\delta,\lambda)} \leq \frac{6+3\gamma}{10+9\gamma}$. Therefore, the $\ell_m^*(\delta,\lambda,\theta) < \epsilon_s^*(\delta,\lambda)$ provided $\theta \geq \left(\frac{10+9\gamma}{4+6\gamma}\right)\lambda$.

S1.2 Diffusion-advection mobile architectures

The perfectly mobile architecture requires a perfectly co-ordinated mechanism to transport sensors, and, in addition, the signal sampling time has to be perfectly co-ordinated with the movement when the radius

	$-1 < \gamma < 0$	$\gamma = 0$	$\gamma > 0$
$\lambda \leq \sqrt{(1+\gamma)\delta}$	Stationary	Stationary = Mobile	Mobile
$\sqrt{(1+\gamma)\delta} < \lambda \leq 3\sqrt{(1+\gamma)\delta}$	Stationary $\to$ Mobile	Mobile	Mobile
$3\sqrt{(1+\gamma)\delta} < \lambda$	Mobile	Mobile	Mobile

Table 1: Optimal architecture as function of $(\delta, \lambda, \gamma)$

$r^* > vt_m$. Next, we describe a architecture where the sensors move in an asynchronous manner; however, they are assembled at the new positions by time t_m, and therefore, one does not have to co-ordinate the sampling time. We will call this mechanism the diffusion-advection driven architecture. This architecture is a close approximation for the signaling platforms driven architecture that we describe in the next section.

A fraction κ of all the sensors are assembled at the Voronoi center and the remaining $(1-\kappa)$ fraction of the sensors are freely diffusing in the region. The freely moving sensors are attracted to the new Voronoi centers via an advection mechanism and the uniform density is restored by sensors diffusing from the previous Voronoi sensors. Let t_a denote the advection time required for the sensors move from one hexagonal configuration to another. Since $(1-\kappa)\rho\pi(vt_d)^2$ sensors are assembled at the new Voronoi center, it follows that the fraction of bound sensors κ is the solution of the equation $(1-\kappa)\rho\pi(vt_a)^2 = \frac{3}{2}\kappa\rho r^2$, i.e. $\kappa = \frac{1}{1+\frac{3r^2}{2\pi(vt_a)^2}}$. Note that we are implicitly assuming that only diffusing sensors can be focused at the new centers. Assuming only the sensors at the Voronoi centers are able to estimate the function, the error of this architecture is given by

$$\epsilon_d(\ell; \delta, \lambda, v) = \frac{\ell^2}{3} + \frac{3\delta\lambda^{2(1+\gamma)}}{(1-3\tau_a/\tau)} \cdot \left(1 + \frac{3\ell^2}{2\pi(v\tau_a/\xi)^2}\right)^{(1+\gamma)} \cdot \frac{1}{\ell^{2(1+\gamma)}}$$

We expect $v\tau_a \approx r$, i.e. the advection current collects all the sensors in a radius r around the new Voronoi center. In this case,

$$\epsilon_d(\ell; \delta, \lambda, \theta) = \frac{\ell^2}{3} + \left(1 + \frac{3}{2\pi}\right)^{(1+\gamma)} \frac{\delta\lambda^{2(1+\gamma)}}{(1-\ell/\theta)} \cdot \frac{1}{\ell^{2(1+\gamma)}}$$

From analogy to the perfect mobile architecture,

$$\epsilon_d^*(\delta, \lambda, \theta) = \epsilon_m^*\left(\delta, \left(1 + \frac{3}{2\pi}\right)^{\frac{1}{2}} \lambda, \theta\right).$$

Thus, we see that the threshold value of γ at which the limiting diffusion-advection mechanism is superior to the stationary mechanism is higher than the threshold value of γ for the perfect mobile architecture.

S2 "Realistic" stochastic model: simulation details

Target signal class. The signal $f(\mathbf{x}, t)$ to be estimated is a Gaussian random field with the variance $\mathbf{var}(f(x,t)) = \sigma_f^2$ correlation

$$\mathbf{cor}\,(f(x,t), f(y,s)) = \frac{1}{1 + \frac{\|x-y\|^2}{\xi^2} + \frac{|t-u|^2}{\tau^2}},$$

where ξ denotes the normalized correlation length in space and τ is the correlation length in time. Without loss of generality we will assume that $\mathbb{E}[f(x)] = 0$ for all $x \in \mathcal{L}$.

Sensor movement. The details of the model are as follows. The sensors move on $L \times L$ lattice $\mathcal{L}$ with $L = 2^{11} = 2048$ with toroidal boundary conditions. We considered sensor density ranging from $\rho = 10^{-6}$ to $\rho = 10$, which translates to approximately $N_p = 4$ to $10L^2$ sensors. In the absence of SPs, in each time step each sensor randomly diffuses to one of its 4 nearest neighbors with equal probability. Thus, the diffusion constant D is the same in each direction, and is given by $D = \frac{1}{8}$. The presence of SPs modulates the sensor movement as follows. Let $r(x)$ denote the distance of the lattice position x to the nearest SP location. We assume that the "field" associated with a given SP extends to a radius R. When $r(x) > R$, all sensors at position x are inactive and diffusing. When $r(x) \leq R$, an inactive sensor at location x becomes active with probability p_a. An active sensor at location x is actively transported by the "field" associated with the SP. The potential corresponding to this "field" is given by $V(x) = \beta \min\left\{\frac{r(x)}{R} - 1, 0\right\}$. When $r(x) > 0$, i.e. there is no SP at location x, an active sensor moves to a neighboring lattice location y with probability proportional to $e^{-(V(y)-V(x))}$; consequently, the advection velocity $v = \frac{1}{2}\tanh(\beta/R)$. When $r(x) = 0$, an active sensor does not move. We set $R = 10$ for all our simulations; therefore, the Péclet number $Pe = vR/D = 4R\tanh(\beta/R) = 40\tanh(\beta/10)$. In each time step, each active sensor becomes inactive with probability $1 - p_a$.

Signal estimation. All sensors make a measurement every t_m time units. Therefore, the sensor makes τ/t_m measurements within the correlation time period τ. A sensor at location y at time u is reads

$$f(y,u) + \sigma_p \zeta,$$

where ζ is a standard Normal random variable independent of f. Therefore, if there are $s(y,u)$ sensors co-located at positive y at time u, the estimate

$$\hat{f}(y,u) = f(y,u) + \frac{\sigma_p}{\sqrt{s(y,u)}}\zeta$$

where ζ is a standard Normal random variable independent of f. Note that unlike in the diffusion-advection transport model, *all* sensors sense the signal. The signal $f(x,t)$ is estimated using all measurements $\hat{f}(y,u)$ assuming that the correlation structure of the signal is known. Let $\mathbf{f}$ denote the matrix $\{f(x,t) : x \in \mathcal{L}, t \in [0,T]\}$ stacked up as a column vector, and let $\hat{\mathbf{f}}$ denote the matrix $\{\hat{f}(y,t) : y \in \mathcal{Z}_t, t \in [0,T]\}$, where $\mathcal{Z}_t$ denotes the set of all lattice locations with non-zero sensors. Note that the vector $\hat{\mathbf{f}}$ is random because the sensor locations are random.

Since f is a Gaussian random field, and the noise is also Gaussian, a linear estimator minimizes the mean square error. Let $\mathbf{B}\hat{\mathbf{f}}$ denote a linear estimator for $\mathbf{f}$ given $\hat{\mathbf{f}}$. Then the optimal estimator is given by the solution of the optimization problem

$$\min_{\mathbf{B}} \mathbb{E}\left\|\mathbf{f} - \mathbf{B}\hat{\mathbf{f}}\right\|^2 = \min_{\mathbf{B}} \left\{\mathbb{E}[\|\mathbf{f}\|^2] - 2\,\mathbf{Tr}\left(\mathbf{B}^\top \mathbb{E}[\hat{\mathbf{f}}^\top \mathbf{f}]\right) + \mathbf{Tr}\left(\mathbf{B}^\top \mathbb{E}[\hat{\mathbf{f}}\hat{\mathbf{f}}^\top]\mathbf{B}\right)\right\}$$

Let $\mathbf{B}^*$ denote the optimal solution of this optimization problem. Then the average mean-squared error is given by

$$\epsilon = \frac{1}{T \cdot L^2}\left(\mathbb{E}[\|\mathbf{f}\|^2] - 2\,\mathbf{Tr}\left((\mathbf{B}^*)^\top \mathbb{E}[\hat{\mathbf{f}}^\top \mathbf{f}]\right) + \mathbf{Tr}\left((\mathbf{B}^*)^\top \mathbb{E}[\hat{\mathbf{f}}\hat{\mathbf{f}}^\top]\mathbf{B}^*\right)\right)$$

Note that ϵ is random because the sensor positions are random; however, in the simulation experiments we set T large enough so that ϵ converges to its average value.

Signaling platforms. We assume that n_{sp} SPs are distributed randomly over the $L \times L$ lattice. The average distance between SPs is approximately $\frac{L}{\sqrt{\pi n_{sp}}}$. We assumed that $R = \frac{1}{2} \cdot \frac{L}{\sqrt{\pi n_{sp}}}$. In the rest of this section, we discuss results of our simulations. We simulated the system under three conditions:

(a) Diffusion (Passive diffusion phase): No SPs present and the sensors diffuse in the lattice.

(b) Fixed SPs (Stationary phase): Uniformly placed SPs that do not move and are able to attract the sensors.

(c) Mobile SPs (Active clustering phase): SP locations are chosen uniformly, and SP relocates after an exponentially distributed lifetime τ. Note that the mean SP lifetime is taken to be the same as the signal correlation length τ in time. We are implicitly assuming that the SP lifetime has evolved to "match" the signal correlation in time.

Simulation parameters and error comparison. We compared the mean error achieved by these three signaling strategies as a function of changing the following set of parameters:

(i) SP "field" strength β. The default value was $\beta = 10$.

(ii) The number of SPs n_{sp}. The default value was $n_{sp} = 8$.

(iii) The correlation length ξ. The default value was $\xi = \frac{L}{16}$

(iv) The correlation time τ. The default value was $\tau = 100$

(v) The frequency of measurement $\tau_m = \frac{\tau}{t_m}$. The default value was $\tau_m = 10$.

(vi) The density of sensor ρ. The default value for $\rho = 0.1$.

(vii) The activation probability p_a. The default value was $p_a = 0.78$. This default value was the approximate value where the mobile SP architecture achieve minimum estimation error.

The sensor noise variance σ_p^2 and the signal variance σ_f^2 were held constant. We simulated for $N_{ev} = 1000$ aster break up events in each condition.

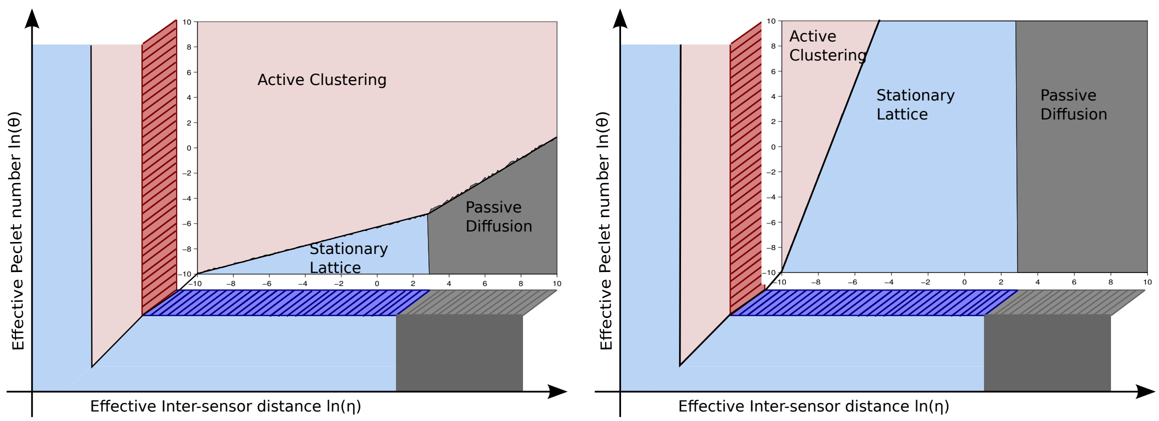
S3 Extension of phase diagram: finite size of sensor

In the simulations (and idealized analysis), the sensors are treated as point particles. However at high enough sensor density or low η, one needs to account for the finite size of sensors. This leads to an extension of the phase diagram as beyond the hashed transition in Fig. 2 of the main manuscript.

The extension of the phase boundary is obtained using the following argument: The number of sensor N at the center of a Voronoi cell with radius r is given by $N = \frac{3\sqrt{3}}{2}\pi\rho r^2$. Suppose each sensor occupies an "effective" area A, where we interpret A to be the minimum area required for effective advection to occur. Suppose the sensors in the center of a Voronoi cell are arranged in a hexagonal close packing with radius R. Then $\frac{3\sqrt{3}}{2}\pi R^2 = NA = \frac{3\sqrt{3}}{2}\pi\rho r^2 A$, i.e. $R = \sqrt{\rho A}r$. In order to differentiate between the active clustering architecture and the stationary architecture, the sensor cluster at the Voronoi cells must not intersect, i.e., $R < r$ or $\rho < 1/A$. Since the effective area A is considerably larger that the sensor area, the upper bound on density is considerably lower than the bound implied by sensors jamming.

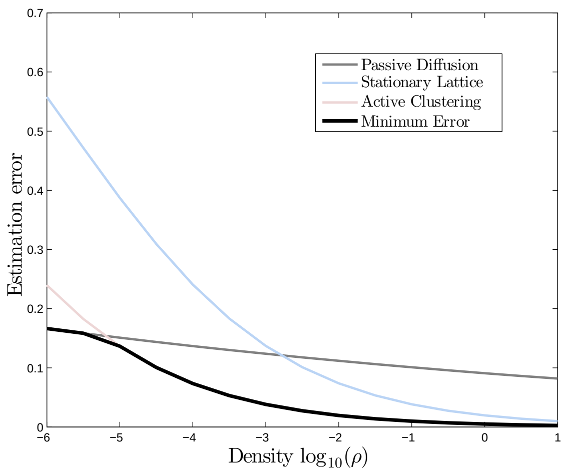
References

1. R. M. Gray, D. L. Neuhoff, *IEEE Trans. Info. Th.* **44**, 2325 (1998).

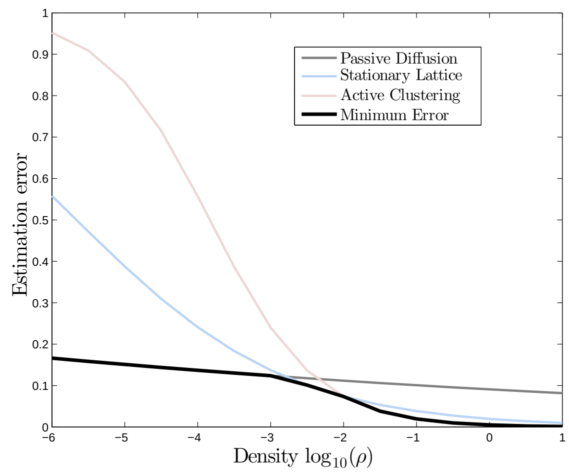


(a) $\gamma = 0.5$ (b) $\gamma = -0.5$

Figure S2: **Phase diagram for correlated sensors in realistic model using Monte Carlo simulations** as a function of (scaled) inter-sensor distance η and Péclet number θ, $\big(\eta \equiv (\sigma_f^2/\sigma_p^2)/(\rho\xi^2)^{\frac{(1+\gamma)}{2}}, \theta \equiv (Pe\,\tau)/\xi\big)$, for (a) Strongly correlated sensors ($\gamma = 0.5$) as a result of allosteric feedback, and (b) Anti-correlated sensors ($\gamma = -0.5$). Phase diagram for independent uncorrelated sensors ($\gamma = 0$) is shown in main text. The phase boundaries are first-order as seen from the intersections of the 'error'-branches as one moves along a cut across the phase diagram (Fig. S3). The measurement frequency $\tau_m = 10$ for all the simulations. Values for the remaining four parameters, Pe, τ, ξ and ρ, were varied within a range and combined to construct the pair (η, θ). In the simulations (and idealized analysis), the sensors are treated as point particles. However at high enough sensor density (low η), one needs to account for steric constraints arising from the finite size of sensors ($S.I$). This leads to an extension of the phase diagram shown beyond the hashed transition boundaries.

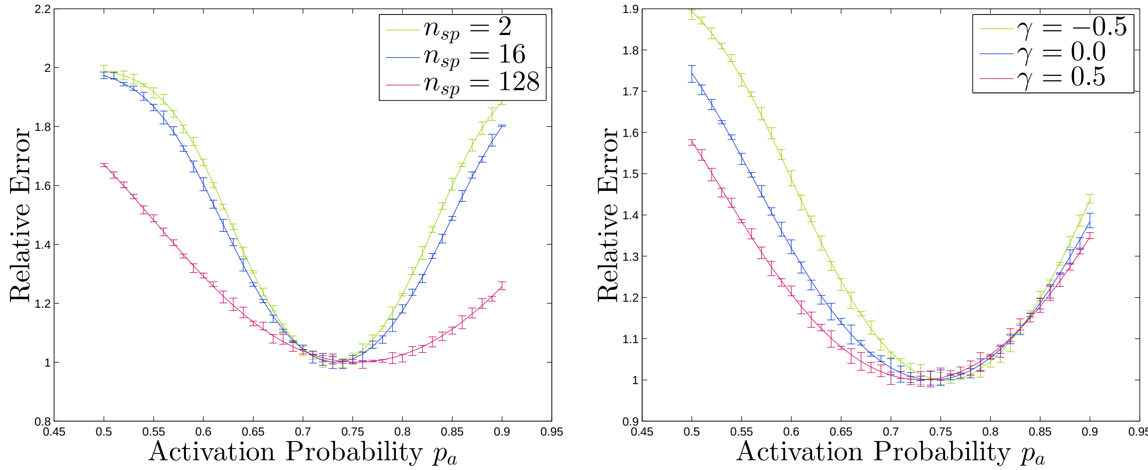


(a) Error vs ρ (high θ)



(b) Error vs ρ (low θ)

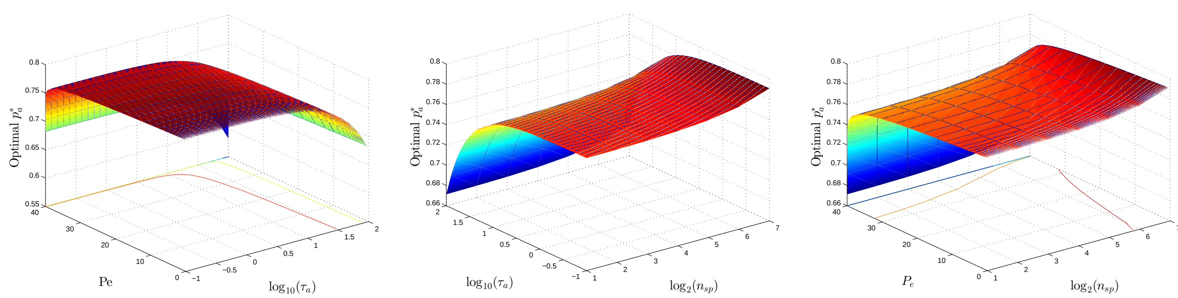
Figure S3: **Intersection of estimation error 'branches'.** (a) Estimation error of the three different phases as a function of ρ for high $\theta = (Pe\ \tau)/\xi$. The optimal phase is that which attains the smallest error for a given density. The optimal architecture goes from being *passive diffusion* to *active clustering* as the sensor density is increased; the stationary lattice is never optimal. (b) Estimation error of the three different phases as a function of ρ for low $\theta = (Pe\ \tau)/\xi$. The optimal architecture goes from being *passive diffusion* to *stationary lattice* to *active clustering* as the sensor density is increased. It is clear from the intersections of the error 'branches', that all the phase transitions are first-order.



(a) Relative error vs n_{sp}

(b) Relative error vs γ

Figure S4: **More on robustness of optimal solution in the active clustering phase**. (a) Plot of relative error in the active clustering phase as a function of activation rate p_a for different values for n_{sp}, the number of SPs. Note that we have normalized the plots for different values of n_{sp} so that the minimum value is 1. (b) Plot of relative error in the active clustering phase as a function of activation rate p_a for different values for sensor correlation γ shows that the minimum is robust at $p_a^* \approx 0.78$, and is more sensitive at low values of γ.



(a) p_a^* vs (τ, Pe)

(b) p_a^* vs (τ, n_{sp})

(c) p_a^* vs (n_{sp}, Pe)

Figure S5: **Surface plots showing robustness of optimal solution in the active clustering phase**. (a)-(c) Optimal probability p_a^* as a function of the pair of parameters $(\cdot, \cdot)$ exhibits a flat unchanged profile over a wide range demonstrating its robustness. Note that except for Pe, the rest of the parameters are varied in log-scale.